CALL FOR PAPERS

Call for Special Issue Papers: Evaluation and Experimental Design in Data Mining and Machine Learning

Deadline for Manuscript Submission: February 1, 2021

Guest Editors: **Eirini Ntoutsi**; *Leibniz University Hannover and L3S Research Center, Germany*;
Erich Schubert; *Technical University Dortmund, Germany*; **Arthur Zimek**; *University of Southern Denmark, Denmark*;
Albrecht Zimmermann; *University Caen Normandy, France*

Special Issue Publication Date: Late 2021

A vital part of proposing new machine learning and data mining approaches is evaluating them empirically to allow an assessment of their capabilities. Numerous choices go into setting up such experiments: how to choose the data, how to preprocess them (or not), potential problems associated with the selection of datasets, what other techniques to compare to (if any), what metrics to evaluate, etc. and last but not least how to present and interpret the results. Learning how to make those choices on-the-job, often by copying the evaluation protocols used in the existing literature, can easily lead to the development of problematic habits. Numerous, albeit scattered, publications have called attention to those questions and have occasionally called into question published results, or the usability of published methods.

Those studies consider different evaluation aspects in isolation, and the issue becomes even more complex because setting up an experiment introduces additional dependencies and biases: having chosen an evaluation metric with little bias can be easily undermined choosing data that cannot appropriately treated by one of the comparison techniques, for instance, and having carefully addressed both aspects is of little worth if the statistical test chosen does not allow to assess significance.

At a time of intense discussions about a reproducibility crisis in natural, social, and life sciences, and conferences such as SIGMOD, KDD, and ECML/PKDD encouraging researchers to make their work as reproducible as possible, we therefore feel that it is important to discuss those issues on a fundamental level. In non-computational sciences, experimental design has been studied in depth, which has given rise to such principles as *randomization*, *blocking*, or *factorial experiments*. While these principles are usually not applied in machine learning and data mining, one desirable goal might be the formulation of a checklist that quickly allows to evaluate the experiment one is about to perform, and to identify and correct weaknesses. An important starting point of any such list has to be: "What question do we want to answer?"

An issue directly related to the dataset choice mentioned above is the following: even the best-designed experiment carries only limited information if the underlying data are lacking. We therefore also want to discuss questions related to the availability of data, whether they are reliable, diverse, and whether they correspond to realistic and/or challenging problem settings. This is of particular importance because our field is at a disadvantage compared to other experimental science: whereas there, data are collected (e.g., in social sciences), or generated (e.g., in physics), we often "only" use existing data.

Finally, we want to emphasize the responsibility of the researchers to communicate their research as objectively as possible. We also want to highlight the critical role of the reviewers: The typical expectation of many reviewers seems to be that an evaluation should demonstrate that a newly proposed method is better than existing work. This can be shown on a

few example datasets at most and is still not necessarily true in general. Rather it should be demonstrated in papers (and appreciated by reviewers) to show on what kind of data a new method works well, and also where it does not, and this way in which respect it is different from existing work and therefore is a useful complement. A related topic is therefore also how to characterize datasets, e.g., in terms of their learning complexity and how to create benchmark datasets, an essential tool for method development and assessment, adopted by other domains like computer vision, IR etc.

Topics for This Special Issue

For this special issue, we mainly solicit contributions that discuss those questions on a fundamental level, take stock of the state-of-the-art, offer theoretical arguments, or take well-argued positions, as well as actual evaluation papers that offer new insights, e.g. question published results, or shine the spotlight on the characteristics of existing benchmark data sets.

As such, topics include, but are not limited to:

- Benchmark datasets for data mining tasks: are they diverse/realistic/challenging?
- Impact of data quality (redundancy, errors, noise, bias, imbalance, ...) on qualitative evaluation
- Propagation/amplification of data quality issues on the data mining results (also interplay between data and algorithms)
- Evaluation of unsupervised data mining (dilemma between novelty and validity)
- Evaluation measures
- (Automatic) data quality evaluation tools: What are the aspects one should check before starting to apply algorithms to given data?
- Issues around runtime evaluation (algorithm vs. implementation, dependency on hardware, algorithm parameters, dataset characteristics)
- Design guidelines for crowd-sourced evaluations
- Principled experimental workflows

Following two workshops on Experimental Design in Data Mining and Machine Learning (EDML) (<https://imada.sdu.dk/Research/EDML/>), we invite now papers on these topics for a special issue of **Big Data**. While extended versions of previous EDML workshop papers are welcome, we also openly invite new submissions that are independent of the EDML workshops.

Please direct special issue inquiries to: Arthur Zimek (zimek@imada.sdu.dk)

Contributions will receive prompt and thorough peer review. Please refer to our Instructions for Authors at www.liebertpub.com/big before submitting your manuscript for consideration.

Big Data is a highly innovative, peer-reviewed journal, provides a unique forum for world-class research exploring the challenges and opportunities in collecting, analyzing, and disseminating vast amounts of data, including data science, big data infrastructure and analytics, and pervasive computing.

Advantages of publishing in *Big Data* include:

- Fast and user-friendly electronic submission
- Rapid, high-quality peer review
- Maximum exposure: accessible in 170 countries worldwide
- Open Access options available

Visit the Instructions for Authors:

www.liebertpub.com/big

Submit your paper for peer review online:

<https://mc.manuscriptcentral.com/big>